Using mixed-effects logistic regression to assess the determinants of regularisation of strong inflection in Dutch

Freek Van de Velde
Britta Kestemont

FWO-Vlaanderen en KU Leuven
Quantitative Lexicology and Variational Linguistics

Introduction

- Strong verbs: seven Germanic ablaut classes

	Infinitive	Preterite sg.	Preterite pl.	Past participle	
Class I	beitan	bait	bitum	bitans	'bite'
Class II	biudan	bauþ	budum	budans	'offer'
Class III	bindan	band	bundum	bundans	'bind'
Class IV	stilan	stal	stelum	stulans	'steal'
Class V	giban	gaf	gebum	gibans	'give'
Class VI	dragan	drog	drogum	dragans	'carry'
Class VII	letan	lailot	lailotum	letans	'let'

- Obscured by:
  - consecutive sound laws in the history of Dutch
  - analogy and contamination

- But several patterns are still recognisable (more so than in English)

Introduction

- Strong verbs: major ablaut classes in Dutch

(see also Knooihuizen & Strik 2014, though note: class III-VII)

	Infinitive	Vowel	Pret. sg.	Vowel	Preterite pl.	Pst. ptcp.	Vowel	
Class I	bijten	/ɛɪ/	beet	/e/	beten	gebeten	/e/	'bite'
Class II	bieden	/i/	bood	/o/	boden	geboden	/o/	'offer'
	zuigen	/œy/	zoog	/o/	zogen	gezogen	/o/	'suck'
Class III	binden	/ɪ/	bond	/ɔ/	bonden	gebonden	/ɔ/	'bind'
	schenden	/ɛ/	schond	/ɔ/	schonden	geschonden	/ɔ/	'violate'
Class IV	stelen	/e/	stal	/ɑ/	stalen	gestolen	/o/	'steal'
Class V	geven	/e/	gaf	/ɑ/	gaven	gegeven	/e/	'give'
	zitten	/ɪ/	zat	/ɑ/	zaten	gezeten	/e/	'sit'
Class VI	dragen	/a/	droeg	/u/	droegen	gedragen	/a/	'carry'
Class VII	laten	...	liet	/i/	lieten	gelaten	...	'let'
Class III-VII	helpen	/ɛ/	hielp	/i/	hielpen	geholpen	/ɔ/	'help'

Introduction

- Strong verbs: major ablaut classes in Dutch

(see also Knooihuizen & Strik 2014, though note class III-VII)

	Infinitive	Vowel	Pret. sg.	Vowel	Pst. ptcp.	Vowel	Pattern	Class size (types)
Class I	bijten	/ɛɪ/	beet	/e/	gebeten	/e/	ABB	55
Class II	bieden	/i/	bood	/o/	geboden	/o/	ABB	13
	zuigen	/œy/	zoog	/o/	gezogen	/o/	ABB	24
Class III	binden	/ɪ/	bond	/ɔ/	gebonden	/ɔ/	ABB	26
	schenden	/ɛ/	schond	/ɔ/	geschonden	/ɔ/	ABB	20
Class IV	stelen	/e/	stal	/ɑ/	gestolen	/o/	ABC	6
Class V	geven	/e/	gaf	/ɑ/	gegeven	/e/	ABA	8
	zitten	/ɪ/	zat	/ɑ/	gezeten	/e/	ABC	3
Class VI	dragen	/a/	droeg	/u/	gedragen	/a/	ABA	5
Class VII	laten	...	liet	/i/	gelaten	...	ABA	20
Class III-VII	helpen	/ɛ/	hielp	/i/	geholpen	/ɔ/	ABC	7

Introduction

- Weakening of strong verbs
  - English: Lieberman et al. (2007), among many others
  - German: Carroll et al. (2012), among many others
  - Frisian: Versloot (2012)
  - Dutch: Van Haeringen (1940), among many others

- Yes, we know the situation is more complicated:
  - weak to strong
  - strong to strong
  - partial weakening
  - back-and-forth switches (strong – weak – strong)
  - 'lectal' variation

Introduction

- Weakening in Dutch. Effect of:
  - **Token frequency** (Van Santen 1997; Vosters 2008, 2012): higher frequency, higher resistance to weakening
  - **Class membership** (van Haeringen, 1940; Van Loey, 1970; Knooihuizen & Strik, 2014): big classes offer protection to weakening
  - **Ablaut pattern** (Van Haeringen 1940; Van Santen 1997): ABA patterns may accelerate weakening of the preterite
  - **Age**: acquisition (Schaerlaekens & Gillis 1987:143-146, 164; Van Santen & Lalleman 1994), apparent time design in Van Santen (1997)

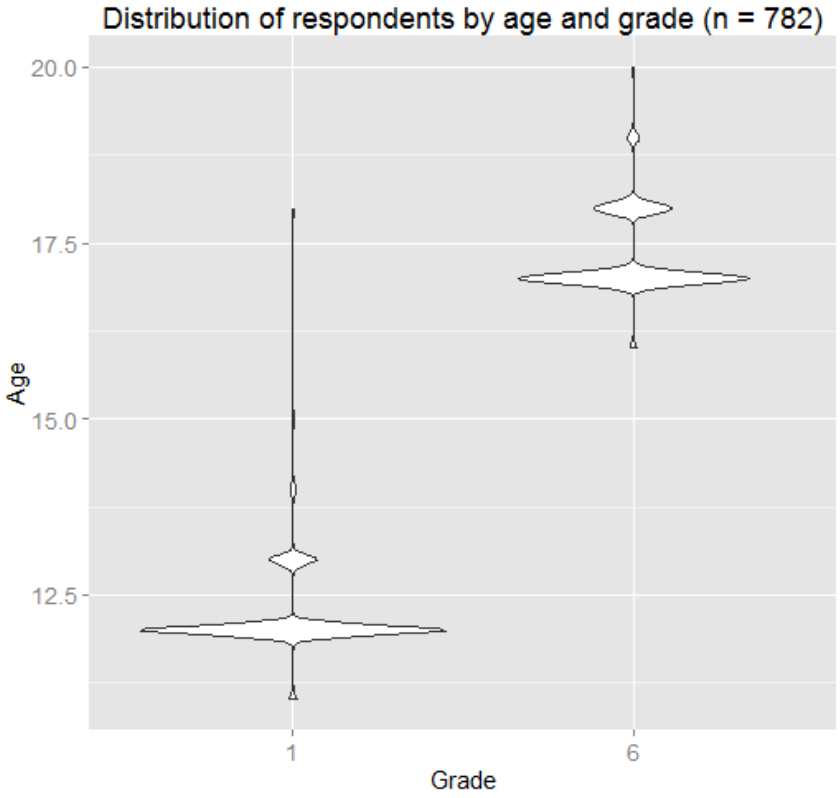
Research question

- What is the effect of token frequency, class membership (size in types, ablaut pattern), and age of the language user on the regularisation process of Dutch verbs when controlling for other variables at issue?

- Design: survey, multivariate analysis (mixed-effect logistic regression)

Survey

- 782 respondents
  - 6 Flemish schools
  - 1st grade secondary education: 415
  - 6th grade secondary education: 367



Survey

- Selection of strong verbs:
  - *Algemene Nederlandse Spraakkunst* (Haeseryn et al. 1997), removed from list:
    - Semi-strong verbs
    - Verbs with vowel and consonant alternation
    - Verbs with 'authorised' weak variant
    - Jocular verbs forms (e.g. *fuiven – foof – gefoven* 'party')
  - Refined by list of strong verbs published in *Onze Taal* (https://onzetaal.nl/taaladvies/advies/sterke-werkwoorden), removed from list:
    - Verbs not present in *Onze Taal* list
    - Verbs with weak variant

Survey

- Classification of selected strong verbs according to their ablaut classes
  - "this categorization has a long tradition and most linguists are familiar with it […] [and] the ablaut classes signify very well the ordering principle behind the strong verbs." (Mailhammer, 2007: 55)

Survey

- Selection and classification selected strong verbs according to their token frequency (Corpus of Spoken Dutch):
  - Lemmatised verbs
  - Four groups based on quartiles
    - High-frequent verbs
    - Semi-high-frequent verbs
    - Semi-low-frequent verbs
    - Low-frequent verbs
  - Dichotomised in the survey
    - frequent
    - infrequent

Survey

- Two filling-in exercises (written)
- Stimulated to fill something in regardless of uncertainty
- Same questionnaire for all respondents

Survey

- First exercise: real verbs
  - Example:

    Voor Nederlands … (VT: lezen) *we vorig jaar "De Aanslag".*
  - 13 weak verbs (fillers)
    - preterite & past participle
    - singular & plural
  - 47 strong verb types
    - preterite & past participle
    - singular & plural

- Second exercise: fake verbs
  - Example: *bleben* ('Class V')

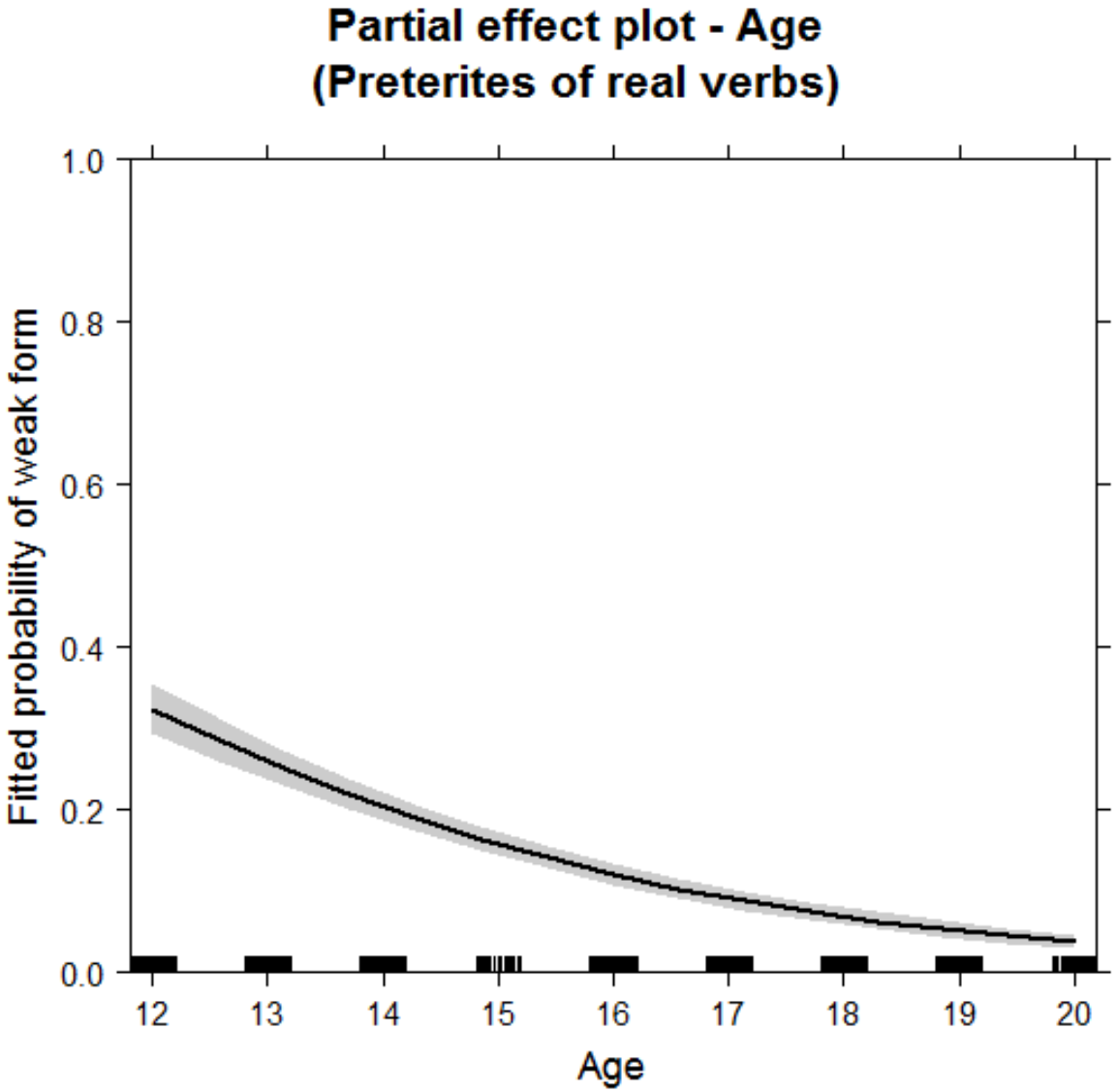
Mixed-effect logistic regression

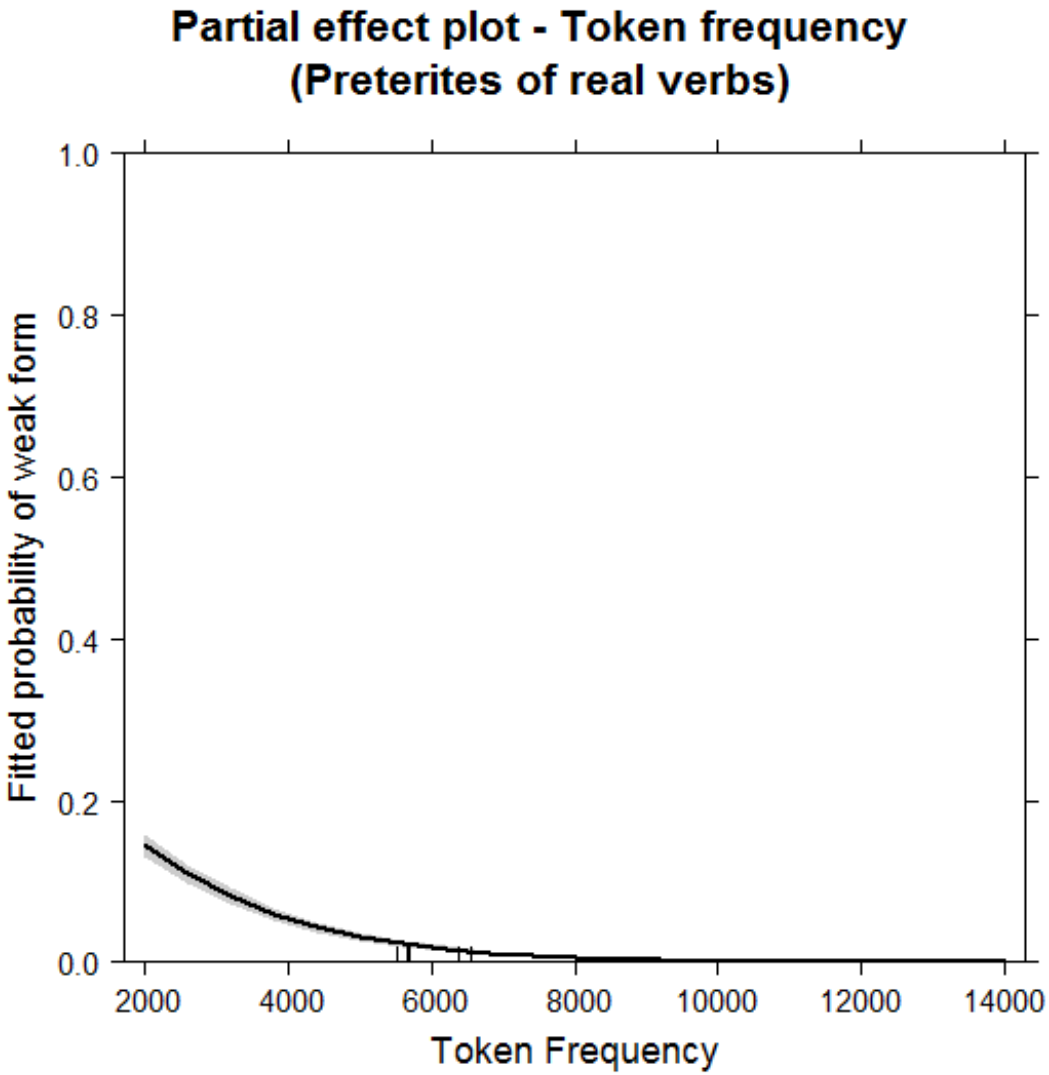
- Response variable:
  - Form (0 strong, 1 weak)
- Fixed effects:
  - Age
  - Token frequency
  - Class size (as ordinal variable or as numeric variable)
  - Pattern (ABB, ABA, ABC)

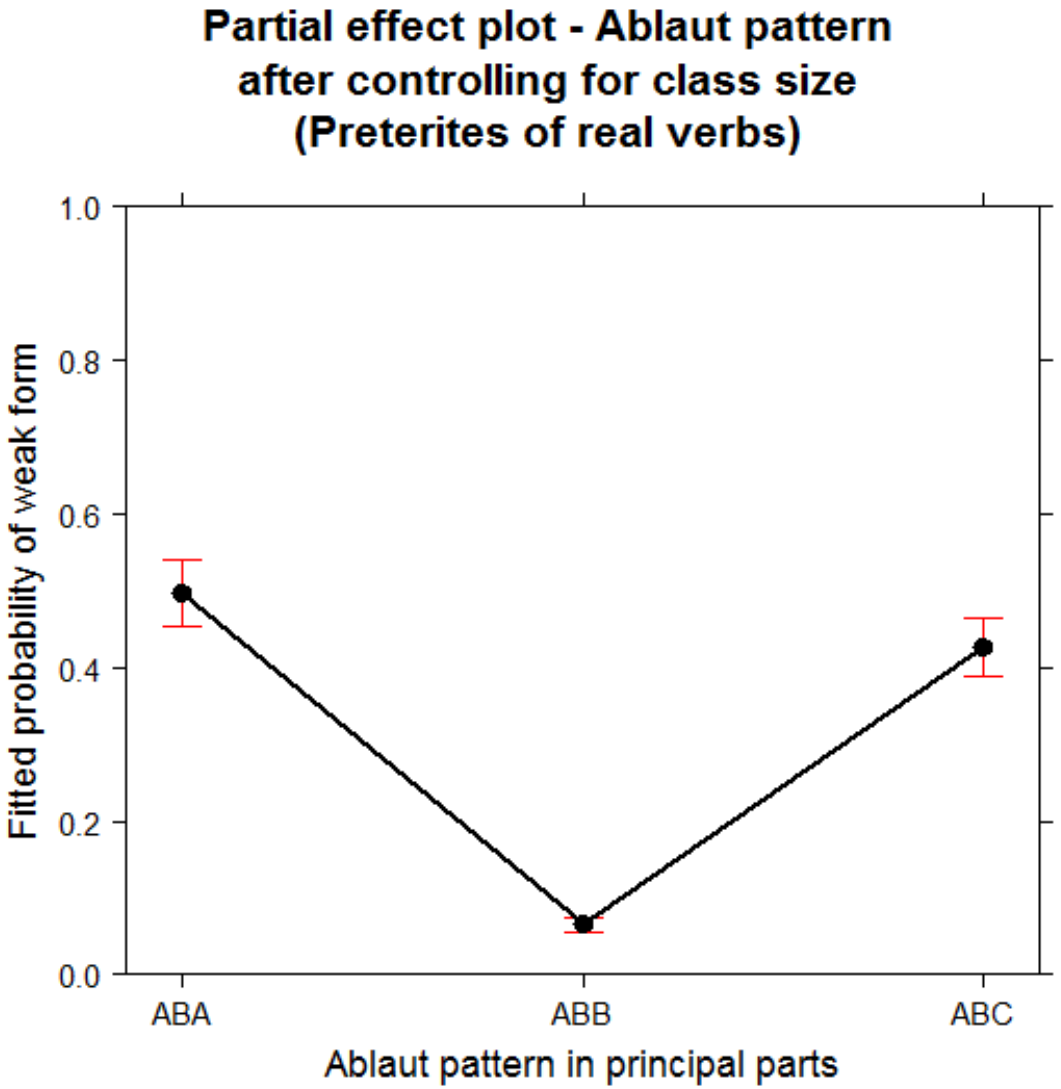
  (no interaction effects, to keep the model simple)
- Random effects:
  - Respondent (remains to be done: nested model with 'school')

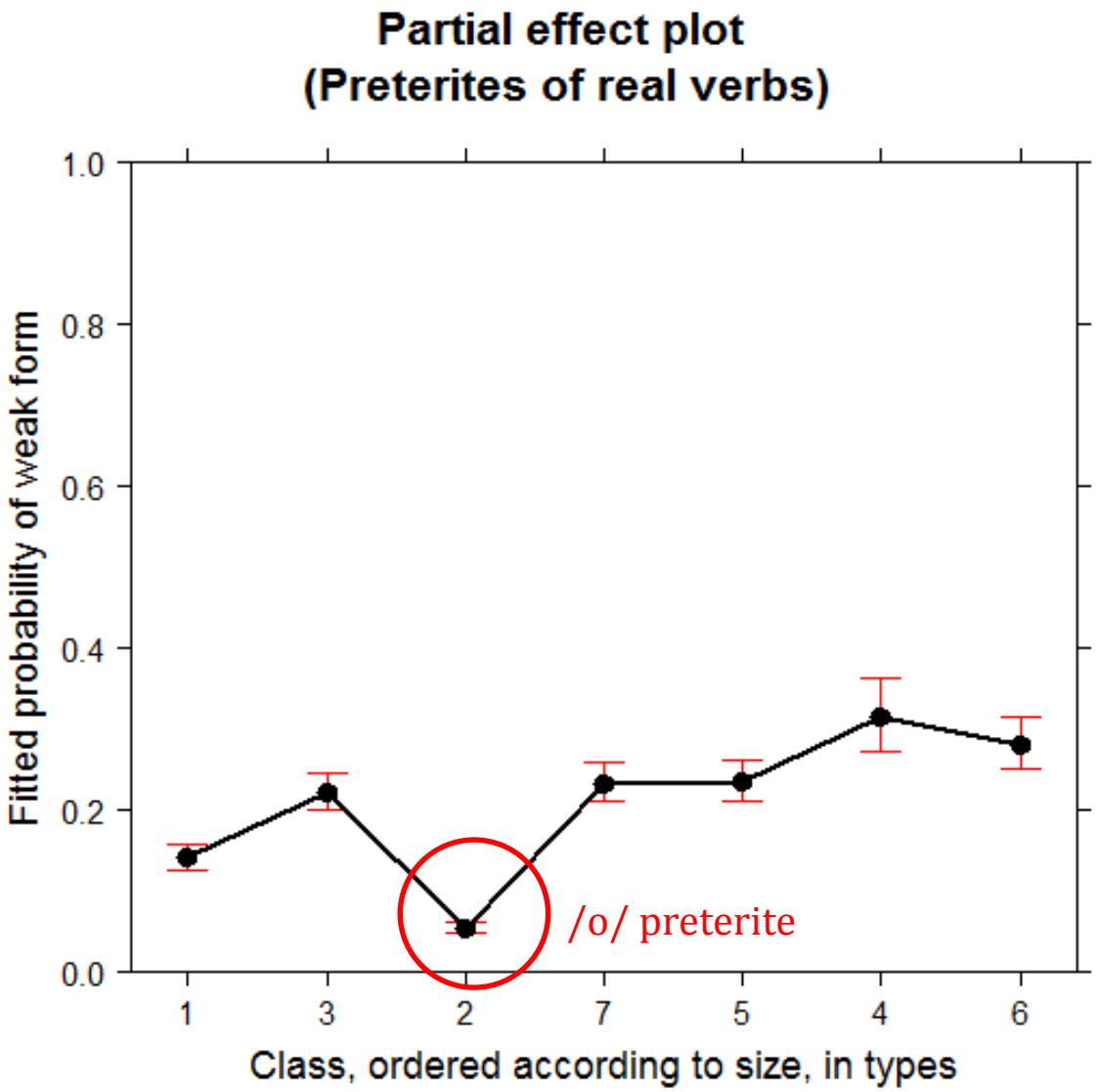
Quality of the models

- All effect plots relate to highly significant predictors ($p < 0.00001$)
- All models have a C value higher than 0.87


Partial effect plot - Age
(Preterites of real verbs)
Fitted probability of weak form
1.0
0.8
0.6
0.4
0.2
0.0
12
13
14
15
16
17
18
19
20
Age


Partial effect plot - Token frequency
(Preterites of real verbs)
Fitted probability of weak form
1.0
0.8
0.6
0.4
0.2
0.0
2000
4000
6000
8000
10000
12000
14000
Token Frequency


Partial effect plot - Ablaut pattern
after controlling for class size
(Preterites of real verbs)
Fitted probability of weak form
1.0
0.8
0.6
0.4
0.2
0.0
ABA
ABB
ABC
Ablaut pattern in principal parts

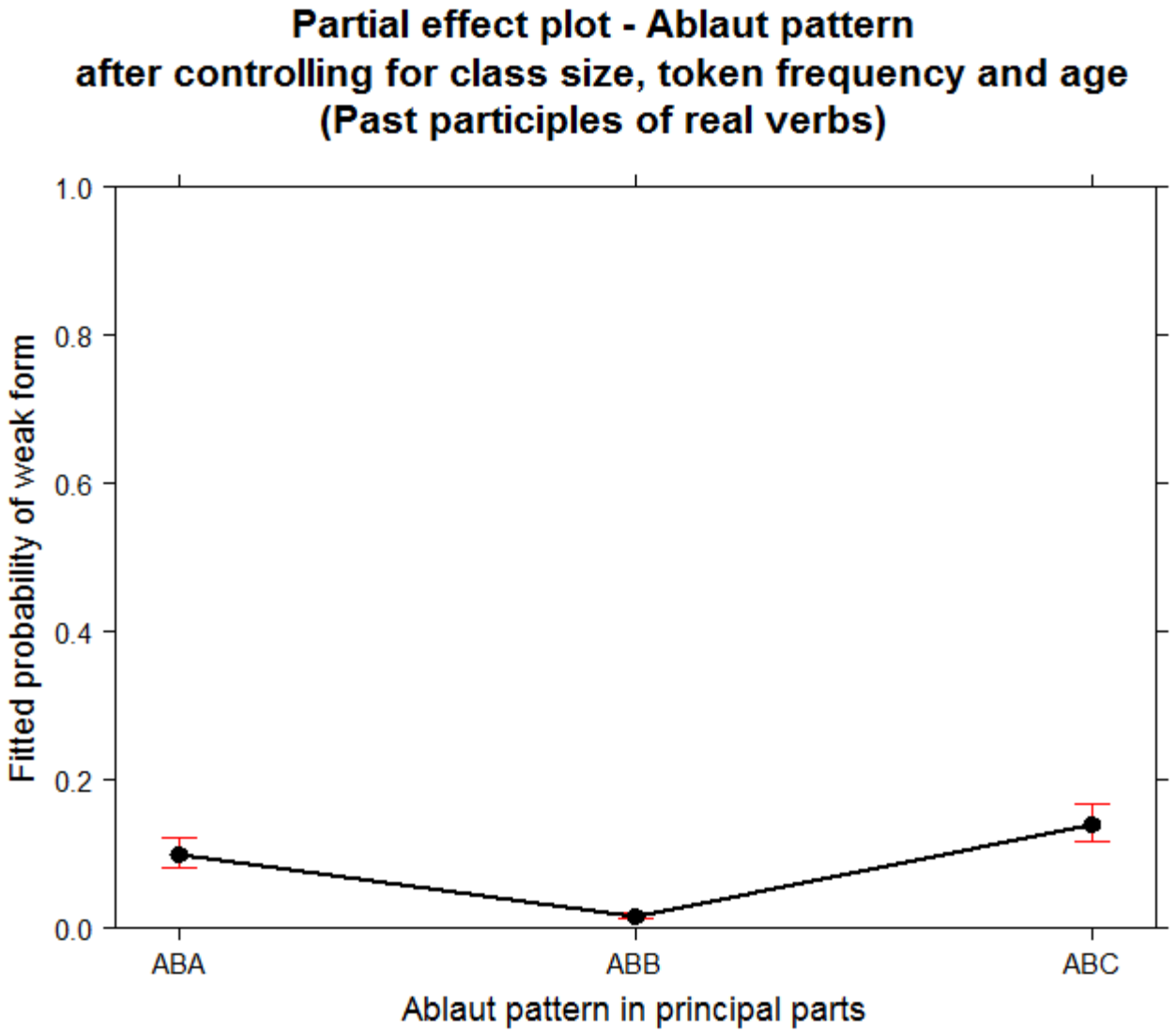

Partial effect plot
(Preterites of real verbs)
Fitted probability of weak form
1.0
0.8
0.6
0.4
0.2
0.0
1
3
2
7
5
4
6
Class, ordered according to size, in types
/o/ preterite

Intermediate conclusions

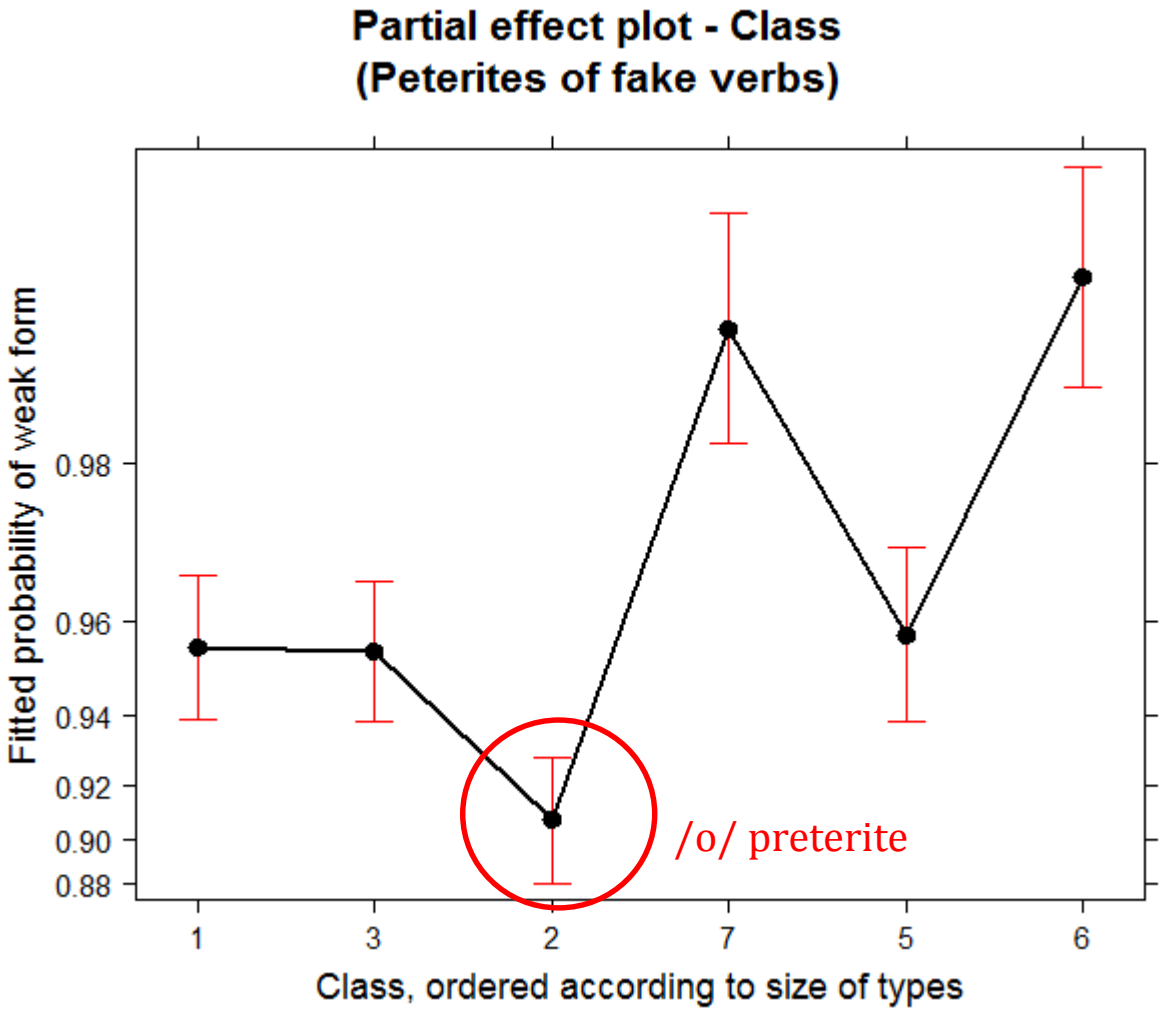
- Multivariate analysis offers confirmation of:
  - Age
  - Token frequency
  - Ablaut pattern
  - /o/ as preterite marker (Knooihuizen & Strik 2014)

- Can we validate these findings?

Validating the ablaut pattern effect



Validating the /o/ effect



(In the fake verbs, Age is no longer a significant predictor)

Assessment

- Shortcomings of previous analyses
  - small-scale
  - no multivariate approach (marginal effects instead of partial effects)

- Shortcomings of our analysis
  - elicited data (underestimation of what young language users do in real life?)
  - small set of verbs (preterite n = 27)
  - synchronic (too short age span for apparent time)

References

- Carroll, R., R. Svare & J. Salmons. 2012. Quantifying the evolutionary dynamics of German verbs. *Journal of Historical Linguistics* 2: 153-172.
- Haeseryn, W., K. Romijn, G. Geerts, J. de Rooij & M.C. van den Toorn. 1997. *Algemene Nederlandse spraakkunst* (2e herziene druk). Groningen/Deurne: Martinus Nijhoff/Wolters Plantyn.
- Knooihuizen, R. & O. Strik. 2014. Relative productivity potentials of Dutch verbal inflection patterns. *Folia Linguistica Historica* 35: 173-200.
- Lieberman, E., J.-B. Michel, J. Jackson, T. Tang & M.A. Nowak. 2007. Quantifying the evolutionary dynamics of language. *Nature* 449: 713-716. doi:10.1038/nature06137.
- Mailhammer, R. 2007. *The Germanic strong verbs: Foundations and development of a new system.* Berlin: Mouton de Gruyter.
- Schaerlaekens, A.M. & S. Gillis. 1987. *De taalverwerving van het kind. Groningen*: Wolters Noordhoff.
- Van Haeringen, C.B. 1940. De taaie levenskracht van het sterke werkwoord. *De Nieuwe Taalgids* 34. 241-255.
- Van Santen, A. 1997. Hoe sterk zijn de sterke werkwoorden?. In: A. Van Santen & M. Van der Wal (eds.), *Taal in tijd en ruimte*. Leiden: SNL, 45-56.
- Van Santen, A. & J. Lalleman. 1994. Gaat zwak anders dan sterk? Over de produktie van Nederlandse regelmatige en onregelmatige verleden-tijdsvormen. *Forum der Letteren* 35(1): 1-22.
- Versloot, A. 2012. *Strong verbs*. Presentation.
- Vosters, R. 2008. Wordt er hier nog Nederlands gespreekt?: een synchronisch onderzoek naar sociale en cognitieve aspecten van werkwoordsverzwakking. *Toegepaste Taalwetenschap in Artikelen* 79: 43-52.
- Vosters, R. 2012. 'Geolinguistic data and the past tense debate. Linguistic and extralinguistic aspects of Dutch verb regularization'. In: G. De Vogelaer & G. Seiler (eds.), *The dialect laboratory. Dialects as a testing ground for theories of language change*. Amsterdam: John Benjamins. 227-248.